arXiv:2304.01134v1 [math.OC] 3 Apr 2023

煤气灯操纵对部分可观测随机控制的影响

Shutian Liu 和 Quanyan Zhu

摘要——近年来，由错误信息导致的网络犯罪和系统故障显著增多。其中许多情况可归类为煤气灯操纵，即通过利用信息操纵他人的感知。本文提出一个基于部分可观测随机控制系统的动态博弈论框架，用于研究煤气灯操纵。博弈中的决策者（DM）只能获取部分观测，并通过构建能够反映其对系统感知的信息状态来确定控制量。博弈中的煤气灯操纵者通过设计观测来操纵决策者的感知和决策，从而间接影响系统。我们通过分析信息状态和最优值对观测偏差的鲁棒性，研究操纵者所作努力的影响。我们引入隐蔽性约束以限制操纵者的能力，并帮助其避免被发现。我们采用近似反馈斯塔克尔伯格均衡作为解的概念，并估算煤气灯操纵成本。

一、引言

煤气灯操纵是一种社会心理虐待，它会操纵人们的信念，使他们质疑自己及其所处的环境。这涉及煤气灯操纵者与受害者之间的认知对抗，而煤气灯操纵者无法直接影响环境。这种对抗类似于决策者（DM）之间对观测的策略性管理以及对感知的操纵。若要成功地对决策者实施煤气灯操纵，煤气灯操纵者必须促使决策者对环境的信念或对感官信息的感知发生变化，这需要对信息进行精密的控制与设计。

信息技术、人工智能（AI）和社交媒体日益普及，使得实施煤气灯操纵的成本更低、更加便利。在网络化信息物理系统中，攻击者可以利用从系统中多个来源收集信息的能力，生成虚假信息并将其传播给用户，逐渐改变他们的感知和操作行为。我们的社会充斥着各种错误信息；如果缺乏检测、辨别和消除错误信息的能力，我们就容易受到煤气灯操纵，被诱导为恶意目的采取行动。因此，深入理解煤气灯操纵者如何设计信息以操纵感知，至关重要。

本文提出一个有限阶段的动态斯塔克尔伯格博弈框架 [1]，用于对煤气灯操纵过程建模。每个阶段都由煤气灯操纵者先行动，随后由决策者行动，这体现了博弈的顺序性：煤气灯操纵者扮演主导角色，而决策者则在不知晓煤气灯操纵者存在的情况下被动地作为跟随者。该动态博弈框架建立在部分可观测随机控制系统之上，其中煤气灯操纵者可以获取状态和观测过程，而决策者只能获取观测过程。煤气灯操纵者掌握完整状态信息，因此占据优势，可以通过设计观测分布来塑造决策者的感知；而决策者的控制决策则取决于其信息状态，这些信息状态通过贝叶斯法则的一个变体生成。煤气灯操纵者通过隐蔽性指标和近似反馈斯塔克尔伯格解来限制其设计力度，以避免被发现；这种解消除了煤气灯操纵力度对状态的直接依赖，从而向决策者隐瞒状态信息。

本文采用由信息状态所定义的 L^1 范数诱导的度量，考察煤气灯操纵的影响，该度量刻画了煤气灯操纵对决策者感知的影响。信息状态对煤气灯操纵力度的鲁棒性用于衡量煤气灯操纵的影响。我们证明，给定观测样本时感知的变化有一个上界，该上界累积了各阶段煤气灯操纵力度的有效作用，与控制无关。我们还研究了煤气灯操纵对决策者所得最优值的影响，从而扩展了鲁棒性分析。

为限制煤气灯操纵者的操纵力度并确保其不被察觉，我们提出一种隐蔽性测度。该测度按阶段定义，以保证在决策者每次获得新的观测并判断其可信度时，煤气灯操纵者都能保持隐蔽。当煤气灯操纵力度具有隐蔽性时，决策者最优值偏差的上界可简化为紧凑形式。这些鲁棒性结果适用于一般的部分可观测随机控制问题。我们采用近似反馈斯塔克尔伯格均衡这一解概念，在其中煤气灯操纵者使用次优的煤气灯操纵力度，从而降低求取精确最优行动的计算负担。这种近似给决策者推断真实状态信息增加了额外难度。我们还估计了煤气灯操纵者在其操纵力度具有隐蔽性时所付出的成本，刻画了煤气灯操纵者试图实施操纵而不被察觉时预期会产生的后果。

我们在第二节回顾相关研究。在介绍作为基础的部分可观测随机控制问题后，第三节将提出用于建模煤气灯操纵的动态斯塔克尔伯格博弈框架。第四节将分析煤气灯操纵的影响并给出煤气灯操纵博弈的解。第五节总结全文。

II. 相关研究

我们的框架建立在部分可观测随机控制问题的经典信息状态方法之上[2]，[3]，[4]，，但采用博弈论视角，并

聚焦于信息设计的影响。与以往将问题表述为小噪声极限或采用 H无穷准则的研究不同，我们采取分布视角，并考虑观测分布的设计。此外，我们框架中的煤气灯操纵者扮演一般性角色，并拥有自己的目标函数需要优化，因此形成了非零和设定。

我们的框架与单边部分可观测随机博弈（OSPOSG）的相关文献密切相关 [5]，[6], [7], [8]。OSPOSG 对现实中的各种策略关系进行建模，并避免了对信念层级的不必要推导。在我们的框架中，煤气灯操纵者的角色有所不同，因为他们无法直接控制博弈状态，而必须通过构造特定的观测分布，影响另一位参与者的行动，进而影响状态演化。这种间接性使得研究煤气灯操纵对系统的影响颇具挑战。受 [9], [10], 的启发，我们通过研究信息状态和最优值对观测偏差的鲁棒性来应对这一挑战，其中观测偏差采用 [2]中提出的度量加以衡量。

我们的工作也为战略性感知操纵这一新兴研究领域作出了贡献。文献 [11] 采用斯塔克尔伯格博弈形式，研究静态情境下的风险感知设计。文献 [12] 通过信息设计研究解除隔离措施后的重新开放政策，文献 [13] 则通过信息设计探讨披露特权知识的潜在优势。在文献 [14], 中，作者还考察了群体决策中的策略性操纵。

III. 问题表述

A. 预备知识

考虑以下离散时间系统：

$$\begin{cases} x_{k+1} = b(x_k, u_k) + w_k, \\ y_{k+1} = h(x_k) + v_k \end{cases} \tag{1}$$

在有限时间区间 $\mathrm{k} = 0, 1, \cdots, \mathrm{K}$ 上的概率空间 $(\Omega, \mathscr{F}, \mathscr{P}^u)$ 中。过程 $\mathrm{x} \in \mathscr{X} \subset \mathbb{R}^n$ 是系统的状态过程，决策者（DM）无法观测该过程。过程 $\mathrm{y} \in \mathscr{Y} \subset \mathbb{R}$ 表示决策者的观测。假设对于所有 $\mathrm{k} = 0, 1, \cdots, \mathrm{K}$，$\mathrm{w}_k \in \mathbb{R}^n$ 均为独立同分布随机变量，其密度函数为 $\psi(\mathrm{w})$。类似地，假设对于所有 $\mathrm{k} = 0, 1, \cdots, \mathrm{K}$，$\mathrm{v}_k \in \mathbb{R}$ 均为独立同分布随机变量，其密度函数为 $\phi(\mathrm{v})$，并且对于所有 $\mathrm{k} = 0, 1, \cdots, \mathrm{K}$，$\mathrm{v}_k \in \mathbb{R}$ 与 x_0 和 w_k 相互独立。阶段 k 的控制 u_k 属于紧集 $\mathrm{U}_k \subset \mathbb{R}^m$。

通过引入以下拉东-尼科迪姆导数，定义概率测度 $\mathscr{P}^\dagger$：

$$\frac{d\mathscr{P}^u}{d\mathscr{P}^\dagger} = Z_k = \prod_{i=1}^{k} \Psi(x_{i-1}, y_i), \tag{2}$$

其中

$$\Psi(x, y) := \frac{\phi(y - h(x))}{\phi(y)}. \tag{3}$$

然后，在测度 $\mathscr{P}^\dagger$ 下，观测 $\mathrm{y1}, \cdots$、y_K 独立同分布，其密度为 ϕ。

决策者的成本函数具有以下形式：

$$J(u) = \mathbb{E}^u \left[\exp\mu \left(\sum_{i=0}^{K-1} L(x_i, u_i) + \Phi(x_K) \right) \right], \tag{4}$$

其中 $\mu > 0$ 表示风险敏感性。

令 $\sigma_k \in \mathrm{L}^1(\mathbb{R}^n)$ 表示第 k 阶段的信息状态。信息状态的更新由有界线性算子 $\Sigma^* : \mathrm{L}^\infty(\mathbb{R}^n) \to \mathrm{L}^\infty(\mathbb{R}^n)$ 控制，其定义如下：

$$\begin{aligned} \Sigma^*(u, y)\sigma(z) = \int_{\mathbb{R}^n} & \psi(z - b(\xi, u)) \\ & \cdot \exp(\mu L(\xi, u)) \Psi(\xi, y)\sigma(\xi) d\xi, \end{aligned} \tag{5}$$

由此得到以下递推，其初始条件为 $\mathscr{X}$ 上的初始概率密度 ρ：

$$\begin{cases} \sigma_k = \Sigma^*(u_{k-1}, y_k)\sigma_{k-1}, \\ \sigma_0 = \rho. \end{cases} \tag{6}$$

为简化记号，定义 T (u,y,z ξ) = ψ(z −b(ξ,u))exp(μL(ξ,u)) ϕ(y −h(ξ))。

损失函数 (4) 可在测度 $\mathscr{P}^\dagger$下表示为：

$$J(u) = \mathbb{E}^\dagger \left[\prod_{i=1}^{K} \Psi(x_{i-1}, y_i) \cdot \exp\mu \left(\sum_{i=0}^{K-1} L(x_i, u_i) + \Phi(x_K) \right) \right], \tag{7}$$

由于应用测度变换 (2) 后，随机变量 $\mathrm{y1}, \cdots, \mathrm{y}_K$ 独立同分布，且其密度为 ϕ，因此可以计算期望 $\mathbb{E}^\dagger[\cdot]$。文献 [3] 证明，(7) 可等价地用信息状态表示如下：

$$\mathscr{J}(u) := \mathbb{E}^\dagger \left[\int_{\mathbb{R}^n} \sigma_K(z) \exp(\mu \Phi(z)) \, dz \right]. \tag{8}$$

B. 煤气灯操纵博弈

本节介绍网络化信息物理系统中决策者与煤气灯操纵者之间的斯塔克尔伯格博弈。决策者根据对系统的局部知识采取动态行动，而煤气灯操纵者则试图操纵决策者的感知，以影响其行动。煤气灯操纵者可以获取系统 x 和 y 的全局信息，而决策者只能观测观测过程 y。这种知识不对称由动力系统 (1) 建模。在该博弈中，决策者作为跟随者，决定系统 (1) 的控制量。其目标是最小化成本函数 (4)。由于决策者缺乏系统真实状态的信息，她会利用观测构建信息状态，以此表示自己对环境的感知。煤气灯操纵者，即领导者，无法直接影响系统演化，但可以通过改变决策者接收到的观测来操纵其感知。因此，他试图通过影响决策者选择的控制量来影响系统。

令 $\mathbb{P}$ 表示$\mathbb{R}$上的概率密度函数集合。煤气灯操纵者的操纵力度由一系列关于观测的概率密度函数 $\phi^\circ = (\phi_1^\circ, \cdots, \phi_K^\circ)$ 表示，即对于所有 $\mathrm{k} = 1, 2, \cdots, \mathrm{K}$，均有 $\phi_k^\circ \in \mathbb{P}$。第 k 阶段的操纵力度$\phi_k^\circ(\mathrm{y}_k)$会影响以下内容的生成：

第 k 阶段的信息状态，它代表决策者对状态 x_k的感知视角。具体而言，在煤气灯操纵力度 $\phi_1^\circ,\cdots,\phi_K^\circ$下，信息状态 $\sigma_1^\circ,\sigma_2^\circ,\cdots,\sigma_K^\circ$的更新变为:

$$\begin{aligned}\sigma_{k+1}^\circ &= \Sigma_k^{*\circ}(u,y)\sigma_k^\circ(z)\\ &= \int_{\mathbb{R}^n}\psi(z-b(\xi,u))\exp(\mu L(\xi,u))\,\Psi_k^\circ(\xi,y)\sigma_k^\circ(\xi)d\xi,\end{aligned} \tag{9}$$

其中 $\Psi_k^\circ(x,y) = \frac{\phi}{(y-h(x))}\phi^\circ{}_{k(y)}$ ·注意，这不会影响测度 $\mathscr{P}^\dagger$。上述设定的含义是，决策者的假设与环境的实际情况并不一致。在决策者的假设中，观测 $y_1, y2,\cdots, y_K$ 是独立同分布的，且服从概率密度 ϕ。因此，她使用 ϕ 来构造概率测度$\mathscr{P}^\dagger$。然而，实际上煤气灯操纵者的确存在，并且决策者的信息状态是根据式（9）在煤气灯操纵力度 ϕ°下生成的。

煤气灯操纵者的目标是在有限阶段内引导系统到达有利状态。我们用函数 $\Gamma:\mathscr{X}\to\mathbb{R}$ 表示终端状态成本，并用函数 $H:\mathbb{P}\to\mathbb{R}_+$ 表示设计成本，其中 $H(\phi)=0$。煤气灯操纵者的问题总结如下:

$$\min_{\phi^\circ}\quad \mathscr{I}(\phi^\circ) := \mathbb{E}^{u^\circ}[\exp(\mu\Gamma(x_K))] - \gamma + \sum_{i=1}^{K} H(\phi_i^\circ), \tag{10}$$

其中， $\mathbb{E}^{u^\circ}$ 表示相对于由决策者的控制 $u^\circ=(u_1^\circ, u_2^\circ,\cdots, u_K^\circ)\in U_1\times U_2\times\cdots\times U_K$所诱导的概率测度的期望算子；当她的观测遵循 ϕ° 和 $\gamma := \mathbb{E}^u[\exp(\mu\Gamma(x_K))]$ 时，u 表示决策者的控制；当她的观测是独立同分布且遵循 ϕ 时，u 也表示决策者的控制。注意，(10) 中的项 γ 是一个归一化常数，因为其值独立于 ϕ°。

决策者针对煤气灯操纵力度ϕ° 作出的响应问题具有如下形式:

$$\min_{u^\circ}\quad \mathscr{J}^\circ(u^\circ) := \mathbb{E}^\dagger\left[\int_{\mathbb{R}^n}\sigma_K^\circ(z)\exp(\mu\Phi(z))\,dz\right], \tag{11}$$

其中，信息状态 σ_1°、σ_2°、 $\cdots$、 σ_K° 根据 (9) 生成。

IV. 煤气灯操纵的影响

本节首先通过分析决策者的信息状态和最优值对煤气灯操纵力度的鲁棒性，探讨煤气灯操纵的影响。在分析过程中，我们对煤气灯操纵力度施加约束，以保证其隐蔽性。当煤气灯操纵具有隐蔽性时，鲁棒性结果的形式较为简单。随后，我们利用鲁棒性结果估计煤气灯操纵者的最优设计成本，其核心是近似反馈斯塔克尔伯格均衡这一解概念。

A. 对煤气灯操纵力度的鲁棒性

分析分为两个步骤。首先，我们校准信息状态过程对于观测过程分布变化的鲁棒性。然后，我们利用这些结果研究最优值对于信息状态过程偏差的鲁棒性。

令 $\sigma_0,\sigma_1,\cdots,\sigma_K$ 表示根据 $\sigma_{k+1}=\Sigma^*(u_k,y_{k+1})\sigma_k$ 更新的原始信息状态过程，其中观测分布遵循 ϕ。令 $\sigma_0^\circ,\sigma_1^\circ,\cdots,\sigma_K^\circ$ 表示根据 $\sigma_{k+1}^\circ=\Sigma_{k+1}^{*\circ}(u_k,y_{k+1})\sigma_k^\circ$ 更新的被操纵的信息状态过程，其中观测分布遵循 $\phi_{k+1}^\circ(y_{k+1})$。我们使用记号 $\sigma^{\phi}{}_{k+1}(\sigma_k)$ 表示阶段 $k+1$的信息状态，该状态通过使用一个涉及 ϕ 的算子对前一信息状态 σ_k 进行更新而得到。

回想一下，信息状态 σ 是未归一化密度函数，即 $\sigma\in L^1(\mathbb{R}^n)$。对于 $\sigma^1,\sigma^2\in L^1(\mathbb{R}^n)$，可以定义由 L^1 范数诱导的度量，形式如下:

$$d(\sigma^1,\sigma^2) = ||\sigma^1-\sigma^2||_{L^1} = \int_{\mathbb{R}^n}|\sigma^1(z)-\sigma^2(z)|dz. \tag{12}$$

a) 信息状态的鲁棒性：令 $\hat{}\phi := \max_y\phi(y)$ 和 $l := \max_{x,u}\exp(\mu L(x,u))$。

引理 1：给定 $\bar{\sigma}_k$、$\hat{}\sigma_k\in L^1(\mathbb{R}^n)$ 和 ϕ，对于给定的观测 y 及所有控制量 u，以下不等式成立:

$$d(\sigma_{k+1}^{\phi}(\bar{\sigma}_k),\sigma_{k+1}^{\phi}(\hat{\sigma}_k)) \le \phi^{-1}(y)\hat{\phi}l\cdot d(\bar{\sigma}_k,\hat{\sigma}_k). \tag{13}$$

证明: 由Fan , we obtain the following equations:

$$\begin{aligned}&d(\sigma_{k+1}^{\phi}(\bar{\sigma}_k),\sigma_{k+1}^{\phi}(\hat{\sigma}_k))\\ =&\int_{\mathbb{R}^n}\left|[\sigma_{k+1}^{\phi}(\bar{\sigma}_k)](z)-[\sigma_{k+1}^{\phi}(\hat{\sigma}_k)](z)\right|dz\\ =&\phi^{-1}(y)\int_{\mathbb{R}^n}\int_{\mathbb{R}^n}T(u,y,z,\xi)\,|\bar{\sigma}_k(\xi)-\hat{\sigma}_k(\xi)|\,d\xi dz\\ =&\phi^{-1}\int_{\mathbb{R}^n}\left(\int_{\mathbb{R}^n}T(u,y,z,\xi)dz\right)\cdot|\bar{\sigma}_k(\xi)-\hat{\sigma}_k(\xi)|\,d\xi\\ =&\phi^{-1}\int_{\mathbb{R}^n}\tilde{T}(u,y,\xi)\cdot|\bar{\sigma}_k(\xi)-\hat{\sigma}_k(\xi)|\,d\xi,\end{aligned} \tag{14}$$

其中

$$\begin{aligned}\tilde{T}(u,y,\xi) &:= \int_{\mathbb{R}^n}T(u,y,z,\xi)dz\\ &= \int_{\mathbb{R}^n}\psi(z-b(\xi,u))dz\cdot\exp(\mu L(\xi,u))\,\phi(y-h(\xi))\\ &= \exp(\mu L(\xi,u))\,\phi(y-h(\xi)).\end{aligned}$$

利用 H¨older 不等式，(14) 得出

$$\begin{aligned}&d(\sigma_{k+1}^{\phi}(\bar{\sigma}_k),\sigma_{k+1}^{\phi}(\hat{\sigma}_k)) \le \phi^{-1}(y)\cdot\\ &\left(\int_{\mathbb{R}^n}|\bar{\sigma}_k(\xi)-\hat{\sigma}_k(\xi)|^p d\xi\right)^{1/p}\cdot\left(\int_{\mathbb{R}^n}|\tilde{T}(u,y,\xi)|^q d\xi\right)^{1/q},\end{aligned} \tag{15}$$

对于满足 $1/p+1/q=1$ 的 $p,q\in[1,\infty]$。取 $p=1$ 和 $q=\infty$，(15) 化简为

$$\begin{aligned}d(\sigma_{k+1}^{\phi}(\bar{\sigma}_k),\sigma_{k+1}^{\phi}(\hat{\sigma}_k)) &\le \phi^{-1}(y)d(\bar{\sigma}_k,\hat{\sigma}_k)\sup_\xi\tilde{T}(u,y,\xi)\\ &\le \phi^{-1}(y)\hat{\phi}l\cdot d(\bar{\sigma}_k,\hat{\sigma}_k).\end{aligned}$$

证明至此完成。 ■

由于对于所有 $k=0,1,\cdots,K$，$\sigma_k\in L^1(\mathbb{R}^n)$，我们用 $\zeta>0$ 表示所有可能信息状态的最大 L^1-范数，即 $\zeta=\max_{k,\sigma_k}||\sigma_k||_{L^1}$。

引理 2：给定 ϕ、ϕ° 和 $\sigma_k \in L^1(\mathbb{R}^n)$，对于给定观测 y 和所有控制量 u，以下不等式成立：

$$d(\sigma_{k+1}^{\phi^\circ}(\sigma_k), \sigma_{k+1}^{\phi}(\sigma_k)) \le \hat{\phi} l\zeta \cdot \left| \frac{1}{\phi^\circ(y)} - \frac{1}{\phi(y)} \right|. \tag{16}$$

因此，我们得到以下方程：

$$\begin{aligned} & d(\sigma_{k+1}^{\phi^\circ}(\sigma_k), \sigma_{k+1}^{\phi}(\sigma_k)) \\ &= \int_{\mathbb{R}^n} \left| [\sigma_{k+1}^{\phi^\circ}(\sigma_k)](z) - [\sigma_{k+1}^{\phi}(\sigma_k)](z) \right| dz \\ &= \left| \frac{1}{\phi^\circ(y)} - \frac{1}{\phi(y)} \right| \cdot \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} |T(u,y,z,\xi)\sigma_k(\xi)| d\xi\, dz \\ &= \left| \frac{1}{\phi^\circ(y)} - \frac{1}{\phi(y)} \right| \cdot \int_{\mathbb{R}^n} \tilde{T}(u,y,\xi)\sigma_k(\xi) d\xi. \end{aligned} \tag{17}$$

由于 $\int_{\mathbb{R}^n} \tilde{T}(u,y,\xi)\sigma_k(\xi) d\xi \le +\infty$ 对所有 u 和 y 均成立，我们得到以下不等式：

$$\begin{aligned} & \int_{\mathbb{R}^n} \tilde{T}(u,y,\xi)\sigma_k(\xi) d\xi \\ &= \int_{\mathbb{R}^n} \exp(\mu L(\xi,u))\, \phi(y - h(\xi))\sigma_k(\xi) d\xi \\ &\le \int_{\mathbb{R}^n} \max_u \exp(\mu L(\xi,u)) \max_y \phi(y - h(\xi))\sigma_k(\xi) d\xi \\ &\le \hat{\phi} l\zeta. \end{aligned} \tag{18}$$

结合 (17) 和 (18) 可得 (16)。 ■

设 $c_k = \left| \frac{1}{\phi^\circ} - \frac{1}{\phi} \right|$ $k = 1, 2, \cdots$

$\hat{l}$。设 $d_k = d(\cdot,\cdot)$。设 $Y_k(y) = \sigma_k(y)$，其中 k = 1, 2, ..., K。煤气灯操纵对决策者感知的影响总结如下。

定理 1：给定观测 $y_1, y_2, \cdots, y_k$，无论控制量 $u_0, u_1, \cdots, u_{k-1}$ 如何，煤气灯操纵力度 $\phi_1^\circ, \phi_2^\circ, \cdots, \phi_k^\circ$ 下阶段 k 的信息状态偏差均满足以下不等式：

$$\begin{aligned} d_k \le & \frac{c^k d_0}{\phi(y_k)\cdot\cdots\cdot\phi(y_1)} + \frac{c^k \zeta Y_1(y_1)}{\phi(y_k)\cdot\cdots\cdot\phi(y_2)} \\ & + \frac{c^{k-1}\zeta Y_2(y_2)}{\phi(y_k)\cdot\cdots\cdot\phi(y_3)} + \cdots + \frac{c^2 \zeta Y_{k-1}(y_{k-1})}{\phi(y_k)} + c\zeta Y_k(y_k). \end{aligned} \tag{19}$$

证明：首先利用三角不等式，观察到以下关系：

$$\begin{aligned} d_{k+1} =& d(\sigma_{k+1}^{\phi_{k+1}^\circ}(\sigma_k^\circ), \sigma_{k+1}^{\phi}(\sigma_k)) \\ \le & d(\sigma_{k+1}^{\phi_{k+1}^\circ}(\sigma_k^\circ), \sigma_{k+1}^{\phi}(\sigma_k^\circ)) + d(\sigma_{k+1}^{\phi}(\sigma_k^\circ), \sigma_{k+1}^{\phi}(\sigma_k)). \end{aligned}$$

然后，引理 1 和引理 2 中的断言可得

$$d_{k+1} \le \frac{c d_k}{\phi_{k+1}(y_{k+1})} + c\zeta \left| \frac{1}{\phi_{k+1}^\circ(y_{k+1})} - \frac{1}{\phi(y_{k+1})} \right|. \tag{20}$$

对式 (20) 使用归纳法可得式 (19)。 ■

b) 煤气灯操纵力度的隐蔽性：我们通过提出以下隐蔽性定义来限制煤气灯操纵者的能力。

定义 1：[期望逐阶段隐蔽性（ESS）]煤气灯操纵力度 ϕ_k° 在阶段 k = 1、2、⋯、K 时为 s-ESS，

如果以下不等式成立：

$$\sup_{u_{k-1}\in U_{k-1}, \sigma_{k-1}\in L^1(\mathbb{R}^n)} \mathbb{E}_{\phi(y_k)} \left[d(\sigma_k^{\phi_k^\circ}(\sigma_{k-1}), \sigma_k^{\phi}(\sigma_{k-1})) \right] \le s, \tag{21}$$

其中，$s > 0$ 表示决策者的信任水平。

上述隐蔽性概念实现了以下两个目标。首先，我们希望得到一个条件，使煤气灯操纵力度在平均意义上具有隐蔽性，而非仅针对特定观测具有隐蔽性。其次，由于感知会在每个阶段更新，隐蔽性必须持续贯穿全部 K 个阶段。注意，在式 (21) 中，期望是相对于分布 ϕ 而非 ϕ_k° 取的。这意味着隐蔽性是从决策者的视角而非煤气灯操纵者的视角进行评估的。毕竟，试图检测煤气灯操纵者是否存在的是决策者。

下面的引理给出了 ESS 的一个充分条件。我们将利用该条件简化最优值鲁棒性的结果。

引理 3：若以下条件对 $\bar{s} = s/(c\zeta)$ 成立，则煤气灯操纵力度 ϕ_k° 在阶段 k 是 s-ESS，即阶段 k 满足 (21)：

$$\int_{\mathbb{R}} \left| \frac{\phi(y)}{\phi_k^\circ(y)} - 1 \right| dy \le \bar{s}. \tag{22}$$

证明：该结论由以下不等式得出：

$$\begin{aligned} & \sup_{\substack{u_{k-1}\in U_{k-1}, \\ \sigma_{k-1}\in L^1(\mathbb{R}^n)}} \mathbb{E}_{\phi(y_k)} \left[d(\sigma_k^{\phi_k^\circ}(\sigma_{k-1}), \sigma_k^{\phi}(\sigma_{k-1})) \right] \\ =& \sup_{\substack{u_{k-1}\in U_{k-1}, \\ \sigma_{k-1}\in L^1(\mathbb{R}^n)}} \int_{\mathbb{R}} \phi(y) \left| \frac{1}{\phi_k^\circ(y)} - \frac{1}{\phi(y)} \right| \int_{\mathbb{R}^n} \tilde{T}(u,y,\xi)\sigma_k(\xi) d\xi\, dy \\ \le & \int_{\mathbb{R}} \phi(y) \left| \frac{1}{\phi_k^\circ(y)} - \frac{1}{\phi(y)} \right| dy \\ & \cdot \sup_{\substack{u_{k-1}\in U_{k-1}, \\ \sigma_{k-1}\in L^1(\mathbb{R}^n)}} \int_{\mathbb{R}^n} \exp(\mu L(\xi,u))\, \hat{\phi}\sigma_k(\xi) d\xi \\ \le & c\zeta \int_{\mathbb{R}} \phi(y) \left| \frac{1}{\phi_k^\circ(y)} - \frac{1}{\phi(y)} \right| dy. \end{aligned}$$

■

条件 (21) 显而易见，而充分条件 (22) 则直接将 ESS 与煤气灯操纵力度 ϕ° 联系起来。具体而言，(22) 的左侧衡量了概率密度函数 ϕ 与 ϕ_k° 之间的偏差。

请注意，定义 1 中的隐蔽性水平可以随阶段而变化。在接下来的章节中，为简化起见，我们假设所有阶段 k = 1, 2, ⋯, K 的隐蔽性水平均相同。只需对记号稍作调整，即可将分析扩展至隐蔽性水平随阶段变化的情形。

c) 最优值的鲁棒性：回顾一下，决策者的目标函数为

$$\mathscr{J}_{\sigma_K}(u) := \mathbb{E}^\dagger \left[\int_{\mathbb{R}^n} \sigma_K(z) \exp(\mu\Phi(z))\, dz \right],$$

其中我们使用记号 $\mathscr{J}_{\sigma_K}$，以强调该成本与信息状态过程相关

$\sigma_0, \sigma_1, \cdots, \sigma_K$。我们关注最优目标值对信息状态过程偏差的鲁棒性，即 $\mathscr{J}_{\sigma_K^\circ} - \mathscr{J}_{\sigma_K}$ 的取值。

令 $e_\Phi = \max_{x \in \mathscr{X}} \exp(\mu \Phi(x))$。令 $\tilde{d}_k(y_1, \cdots, y_k)$ 表示（19）的右端。煤气灯操纵对决策者最优值的影响概括如下。

定理2：无论控制量 $u_0, u1, \cdots, u_{k-1}$ 如何，决策者在煤气灯操纵力度 $\phi_1^\circ, \cdots, \phi_K^\circ$ 下的目标值偏差均满足以下不等式：

$$\mathscr{J}_{\sigma_K^\circ} - \mathscr{J}_{\sigma_K} \le e_\Phi \cdot \mathbb{E}^\dagger \left[\tilde{d}_K(y_1, \cdots, y_K) \right]. \tag{23}$$

证明：根据决策者的目标函数，我们可以得到

$$\begin{aligned} \mathscr{J}_{\sigma_K^\circ} - \mathscr{J}_{\sigma_K} =& \mathbb{E}^\dagger \left[\int_{\mathbb{R}^n} (\sigma_K^\circ(z) - \sigma_K(z)) \exp(\mu \Phi(z)) \, dz \right] \\ =& \int_{\mathbb{R}^n} \mathbb{E}^\dagger \left[\sigma_K^\circ(z) - \sigma_K(z) \right] \exp(\mu \Phi(z)) \, dz. \end{aligned} \tag{24}$$

利用 H¨older 不等式，由（24）可得以下不等式：

$$\begin{aligned} & \mathscr{J}_{\sigma_K^\circ} - \mathscr{J}_{\sigma_K} \\ \le & \int_{\mathbb{R}^n} \left| \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} \Pi_{i=1}^K \phi(y_i) \left[\sigma_K^\circ(z) - \sigma_K(z) \right] dy_1 \cdots dy_K \right| dz \\ & \cdot \max_z \exp(\mu \Phi(z)) \\ \le & \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} \Pi_{i=1}^K \phi(y_i) \left[\int_{\mathbb{R}^n} |\sigma_K^\circ(z) - \sigma_K(z)| dz \right] dy_1 \cdots dy_K \\ & \cdot \max_z \exp(\mu \Phi(z)) \\ \le & e_\Phi \cdot \mathbb{E}^\dagger [d_K]. \end{aligned} \tag{25}$$

结合（19）与（25），即可得到该定理。 ■

隐蔽的煤气灯操纵努力可表示为以下紧凑形式。

定理3：假设煤气灯操纵力度 $\phi_1^\circ, \cdots, \phi_K^\circ$ 对 $k = 1, 2, \cdots, K$ 满足式 (22)，则可化简为 (23)

$$\mathscr{J}_{\sigma_K^\circ} - \mathscr{J}_{\sigma_K} \le e_\Phi (c^K d_0 + s \sum_{i=1}^{K-1} c^i). \tag{26}$$

证明：对式 (20) 两边分别关于 $\phi(y_1) \cdots \phi(y_K)$ 取期望，可得

$$\begin{aligned} \mathbb{E}^\dagger [d_k] \le & \mathbb{E}^\dagger \left[\frac{c d_{k-1}}{\phi_k(y_k)} + c\zeta \left| \frac{1}{\phi_k^\circ(y_k)} - \frac{1}{\phi(y_k)} \right| \right] \\ \le & \mathbb{E}_{\phi(y_k)} \left[\frac{c}{\phi(y_k)} \right] \cdot \mathbb{E}^\dagger [d_{k-1}] \\ & + c\zeta \mathbb{E}_{\phi(y_k)} \left[\left| \frac{1}{\phi_k^\circ(y_k)} - \frac{1}{\phi(y_k)} \right| \right] \\ \le & c \cdot \mathbb{E}^\dagger [d_{k-1}] + s, \end{aligned} \tag{27}$$

其中，最后一个不等式源于式 (22) 在阶段 k 成立这一事实。将对式 (27) 的归纳法与式 (25) 相结合，即得定理。 ■

式（26）中的上界表明了根据式（22）在隐蔽性意义下的煤气灯操纵力度所能达到的最大性能该性能通过决策者最优值的偏差进行评估。从决策者的角度来看，减轻煤气灯操纵影响的一种简单方法是降低信任水平 s。

B. 近似反馈斯塔克尔伯格均衡。

本节基于动态规划过程介绍反馈ε-斯塔克尔伯格均衡这一解概念。随后利用第 IV-A 节中得到的鲁棒性结果来估计煤气灯操纵者的成本。

在 [3], 中，作者表明，利用信息状态递推 (6)，通过最小化等价代价函数 (8)，可以使用动态规划方程对部分观测下的系统 (1) 进行控制。令 Z(σ,k) 表示与式 (8) 和 (6) 相关的值函数。动态规划方程具有以下形式 [3]:

$$\begin{cases} Z(\sigma_k, k) = \inf_{u_k \in U_K} \mathbb{E}^\dagger \left[Z(\Sigma^*(u_k, y_{k+1}) \sigma_k, k+1) \right], \\ Z(\sigma_K, K) = \int_{\mathbb{R}^n} \sigma_K(z) \exp(\mu \Phi(z)) \, dz. \end{cases} \tag{28}$$

在第 III-B 节所构建的煤气灯操纵博弈中，煤气灯操纵者和决策者的阶段代价取决于双方的行动。因此，有必要通过考虑煤气灯操纵博弈中的策略关系，扩展式 (28) 中的公式。为此，我们引入以下函数。假设煤气灯操纵力度 $\phi^\circ = (\phi_1^\circ, \phi_2^\circ, \cdots, \phi_K^\circ)$ 和控制力度 $u^\circ = (u_1^\circ, u_2^\circ, \cdots, u_K^\circ)$ 均已固定。令 V $(\sigma_k^\circ, \phi_k^\circ, u_k^\circ, k)$ 与决策者的控制问题相关联，该问题递归定义如下：

$$\begin{cases} V(\sigma_k^\circ, \phi_k^\circ, u_k^\circ, k) = \\ \quad \mathbb{E}^\dagger \left[V(\Sigma_k^{*\circ}(u_k^\circ, y_{k+1}) \sigma_k^\circ, \phi_{k+1}^\circ, u_{k+1}^\circ, k+1) \right], \\ V(\sigma_K^\circ, \phi_K^\circ, u_K^\circ, K) = \int_{\mathbb{R}^n} \sigma_K^\circ(z) \exp(\mu \Phi(z)) \, dz, \end{cases} \tag{29}$$

其中，信息状态 $\sigma_1^\circ, \sigma_2^\circ, \cdots, \sigma_K^\circ$ 根据式 (9) 生成。令 W $(x_k, \phi_k^\circ, u_k^\circ, k)$ 与煤气灯操纵者的设计问题相关联。由于煤气灯操纵者拥有完全观测，且其目标函数具有可加性，W $(x_k, \phi_k^\circ, u_k^\circ, k)$ 通过以下递推关系递归定义：

$$\begin{cases} W(x_k, \phi_k^\circ, u_k^\circ, k) = H(\phi_k^\circ) \\ \quad + \mathbb{E}^{u^\circ} \left[W(b(x_k, u_k^\circ) + w_k, \phi_{k+1}^\circ, u_{k+1}^\circ, k+1) \right], \\ W(x_K, \phi_K^\circ, u_K^\circ, K) = \exp(\mu \Gamma(x_K)). \end{cases} \tag{30}$$

以下结果是动态规划和斯塔克尔伯格博弈序贯性质的推论。

定理 4：若满足以下条件，则对于给定向量 $\varepsilon = (\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_K)$（其中 $\varepsilon_k > 0, \forall k = 1, 2, \cdots, K$），一对策略 $(\bar{\phi}^\circ, \bar{u}^\circ)$ 构成反馈式 ε-斯塔克尔伯格均衡：

$$W(x_k, \bar{\phi}_k^\circ, \bar{u}_k^\circ, k) \le \min_{\phi_k^\circ, u_k^\circ \in R_k(\phi_k^\circ)} W(x_k, \phi_k^\circ, u_K^\circ, k) + \varepsilon_k, \forall x_k, \forall k, \tag{31}$$

其中，$R_k(\phi_k^\circ)$ 表示为所有 σ_k° 及所有 k 定义的响应集，如下所示：

$$\begin{aligned} R_k(\phi_k^\circ) = & \{ \tilde{u}_k \in U_k : V(\sigma_k^\circ, \phi_k^\circ, \tilde{u}_k, k) \\ & = \min_{u_k \in U_k} V(\sigma_k^\circ, \phi_k^\circ, u_k, k) \}. \end{aligned} \tag{32}$$

近似均衡至少具有以下两项优势。首先，它减轻了计算负担。由于煤气灯操纵力度是一个无限维设计对象，在求解定理 4 中的方程时，可以采用阶梯函数进行近似。其次，它防止了信息泄露。由于煤气灯操纵者能够完整观测系统，决策者可以根据煤气灯操纵者采取的行动推断状态信息。然而，通过求解 (31)，煤气灯操纵者打破了其行动对状态信息的直接依赖。由于我们考虑的是有限阶段动态博弈，决策者难以从煤气灯操纵者的行动中准确推断状态信息。关于近似均衡的更多讨论，请参阅 [5]–[15] 。

以下结果揭示了隐蔽的煤气灯操纵所产生的后果。

定理 5：假设：(i) 设计成本函数选为 $\mathrm{H}(\tilde{\phi}) = \mathrm{t}\int_{\mathbb{R}} \left| \frac{\phi}{\tilde{\phi}}_{(y)(y)} - 1 \right| \mathrm{dy}$，其中 $\mathrm{t} > 0$ 表示设计成本参数；(ii) 对所有 $\mathrm{k} = 1, 2, \cdots, \mathrm{K}$，煤气灯操纵者的努力 ϕ° 严格满足（22）。则煤气灯操纵者的成本满足：

$$\mathscr{J}(\phi^{\circ}) \geq -e_{\Gamma}(\hat{\phi}^{K} d_0 + s \sum_{i=1}^{K-1} \hat{\phi}^{i}) + Kt\bar{s}, \tag{33}$$

其中 $\mathrm{e}_{\Gamma} = \max_{x \in \mathscr{X}} \exp(\mu \Gamma(\mathrm{x})$。

证明：根据 III-A 节，我们知道 DM 的目标函数有两个等价表达式，即 $\mathrm{J} = \mathscr{J}$。然后，根据定理 3 可得以下不等式：

$$\begin{aligned} |J(u^{\circ}) - J(u)| = & \left| \mathbb{E}^{u^{\circ}} \left[\exp\mu \left(\sum_{i=0}^{K-1} L(x_i, u_i^{\circ}) + \Phi(x_K) \right) \right] \right. \\ & \left. - \mathbb{E}^{u} \left[\exp\mu \left(\sum_{i=0}^{K-1} L(x_i, u_i) + \Phi(x_K) \right) \right] \right| \\ \leq & e_{\Phi}(c^{K} d_0 + s \sum_{i=1}^{K-1} c^{i}). \end{aligned} \tag{34}$$

在（34）中对所有 $\mathrm{x} \in \mathrm{X}$ 和所有 $\mathrm{u} \in \mathrm{U}$ 取 $\mathrm{L(x,u)} = 0$，可得

$$\left| \mathbb{E}^{u^{\circ}} [\exp\mu \Phi(x_K)] - \mathbb{E}^{u} [\exp\mu \Phi(x_K)] \right| \leq e_{\Phi}(\hat{\phi}^{K} d_0 + s \sum_{i=1}^{K-1} \hat{\phi}^{i}). \tag{35}$$

注意到，(35) 右侧唯一依赖函数 Φ 的项是 e_{Φ}，因此，直接考虑 (35) 中的函数 Γ 即可得到以下不等式：

$$\left| \mathbb{E}^{u^{\circ}} [\exp\mu \Gamma(x_K)] - \mathbb{E}^{u} [\exp\mu \Gamma(x_K)] \right| \leq e_{\Gamma}(\hat{\phi}^{K} d_0 + s \sum_{i=1}^{K-1} \hat{\phi}^{i}). \tag{36}$$

由此可得定理中的断言。 ■

注意，也可以根据不等式 (36) 推导出 $\mathscr{J}(\phi^{\circ})$ 的上界。然而，下界 (33) 更有意义，原因如下。由于 $\Gamma(\cdot)$ 表示煤气灯操纵者的终端状态成本，因此当最终状态由控制 u° 诱导时，其值低于由 u 诱导时是合理的。因此，$\mathbb{E}^{u^{\circ}}[\exp(\mu\Gamma(\mathrm{x}_K))] \leq \gamma = \mathbb{E}^{u}[\exp(\mu\Gamma(\mathrm{x}_K))]$。于是，(33) 右侧的第二项表示在煤气灯操纵力度满足 (22) 时，煤气灯操纵者的终端状态成本所能获得的最大改善。界 (33) 表明了煤气灯操纵者在最理想情况下可以预期的结果。

V. 结论

本文提出了一种动态斯塔克尔伯格博弈论框架，用于建模煤气灯操纵行为及其影响。该框架以部分可观测随机控制系统为基础，其中信息状态刻画了决策者的感知。我们将框架进一步扩展至部分可观测马尔可夫决策过程，其中的状态转移更加一般化。未来将进一步分析决策者感知和最优值对煤气灯操纵力度的敏感性。通过刻画观测变化时控制动作的变化，敏感性分析能够比鲁棒性结果更详细地解释煤气灯操纵的影响；鲁棒性结果则考察最坏情形下的控制量。

参考文献

[1] T. Bas¸ar 和 G. J. Olsder，《动态非合作博弈论》。SIAM，1998年。[2] R. J. Elliott 和 J. B. Moore，“离散时间部分可观测控制”，载于《微分方程》。Routledge，2017年，第 481–490 页。[3] M. R. James、J. S. Baras 和 R. J. Elliott，“部分可观测离散时间非线性系统的风险敏感控制与动态博弈”，《IEEE自动控制汇刊》，第 39 卷，1994年，第 4 期，第 780–792 页。[4] P. R. Kumar 和 P. Varaiya，《随机系统：估计、辨识与自适应控制》。SIAM，2015年。[5] W. Zheng、T. Jung 和 H. Lin，“单边零和部分可观测随机博弈的斯塔克尔伯格均衡”，Automatica，第 140 卷，页码 110231，2022年。[6] K. Hor´ak、B. Bo˘sansk`y、V. Kova˘r´ık 和 C. Kiekintveld，“求解零和单边部分可观测随机博弈”，人工智能，第 316 卷，页码 103838，2023年。[7] J. H¨orner、D. Rosenberg、E. Solan 和 N. Vieille，“关于具有单方信息的马尔可夫博弈”，《运筹学》，第 58 卷，第 4-part-2 期，第 1107–1115 页，2010年。[8] T. Li 和 Q. Zhu，“双边信息不对称下的信号传递承诺”，arXiv 预印本 arXiv:2212.11446，2022年。[9] C. McDonald 和 S. Yuksel，“部分可观测随机控制中对错误先验的鲁棒性及受控滤波器稳定性”，《SIAM控制与优化杂志》，第 60 卷，2022 年，第 2 期，第 842–870 页。[10] A. D. Kara 、 M. Raginsky 和 S. Y¨uksel，“平均成本最优随机控制中对错误模型的鲁棒性与数据驱动学习”，Automatica，第 139 卷，页码 110179，2022年。[11] S. Liu 和 Q. Zhu，“斯塔克尔伯格风险偏好设计”，arXiv 预印本 arXiv:2206.12938，2022年。[12] ——，“Eproach：用于策略性信息设计以实现负责任的 COVID 重启的群体疫苗接种博弈”，载于 2022年美国控制会议（ACC）。IEEE，2022年，第 568–573 页。[13] K. Paarporn 和 P. N. Brown，“一般乐透博弈中的策略性能力揭示”，arXiv 预印本 arXiv:2211.14907，2022年。[14] Y. Sasaki，“两两比较群体决策中的策略性操纵：博弈论视角”，《欧洲运筹学杂志》，第 304 卷，第 3 期，第 1133–1139 页，2023年。[15] F. Laine、D. Fridovich-Keil、C.-Y. Chiu 和 C. Tomlin，“近似广义反馈纳什均衡的计算”，《SIAM优化杂志》，第 33 卷，第 1 期，第 294–318 页，2023年。